DOCUMENT RESUME

ED 418 661 CE 078 411

AUTHOR Linn, Robert L.; Herman, Joan L.
TITLE A Policymaker's Guide to Standards-Led Assessment.
INSTITUTION National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, CA.; Education Commission of the States, Denver, Colo.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.; National Science Foundation, Washington, D.C.
PUB DATE Feb 97
NOTE 32p.
CONTRACT RED-9452301; R305B960002
AVAILABLE FROM ECS Distribution Center, 707 17th Street, Suite 2700, Denver, CO 80202-3427 (Order No. EE-97-1, $15 plus $3 postage and handling).
PUB TYPE Guides - Non-Classroom (055) -- Reports - Evaluative (142)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Academic Standards; *Educational Assessment; *Educational Objectives; Elementary Secondary Education; Evaluation Criteria; Performance; *State Standards; Student Certification; Student Evaluation

ABSTRACT

States across the United States are setting tough new standards, defining what students should know and be able to do. Unlike more traditional assessments, standards-led assessments are more closely linked to curriculum, incorporate preestablished performance goals, and require students to demonstrate a broad range of problem-solving skills. This document summarizes developments in standards-based assessments and discusses the challenges that policymakers and educators face in establishing such systems. The introductory chapter provides an overview of standards-driven assessments and describes the role of assessment in standards-led reform, the ways in which such assessment differs from traditional assessments, and ways to ensure the quality of the assessment systems. Chapter 2 describes the challenges for standards-led assessment systems, which include building state and local consensus, providing strong standards, aligning standards with assessment and instruction, ensuring accurate measures, defining progress, setting the stakes, including all students, estimating costs, addressing legal challenges, building local capacity, and distinguishing assessments. (Contains 43 endnotes.) (CRS)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 428 449

A Policymaker's Guide To Standards-Led Assessment

CRESST

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

EA 028 411

BEST COPY AVAILABLE

A Policymaker's Guide To Standards-Led Assessment

by
Robert L. Linn
Joan L. Herman

National Center for Research on
Evaluation, Standards and Student Testing
(CRESST)

Education Commission of the States
707 17th Street, Suite 2700
Denver, Colorado 80202-3427
303-299-3600
FAX: 303-296-8332
e-mail: ecs@ecs.org
http://www.ecs.org

February 1997

This guide is issued jointly by the Education Commission of the States (ECS) and the National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

The Education Commission of the States is a nonprofit, nationwide interstate compact formed in 1965 to help governors, state legislators, state education officials and others develop policies to improve the quality of education. It is ECS policy to take affirmative action to prevent discrimination in its policies, programs and employment practices.

The National Center for Research on Evaluation, Standards, and Student Testing conducts research on important topics related to K-12 educational testing. CRESST is funded by the U.S. Department of Education, Office of Educational Research and Improvement.

Copies of this guide are available for $10.00 plus postage and handling from the ECS Distribution Center, 707 17th Street, Suite 2700, Denver, Colorado 80202-3427, 303-299-3692. Ask for No. SI-97-3. ECS accepts prepaid orders, MasterCard, American Express and Visa. All sales are final.

To request permission to excerpt part of this publication either in print or electronically, please write or fax Josie Canales, Education Commission of the States, 707 17th St., Suite 2700, Denver, CO 80202-3427; fax: 303-296-8332.

© Copyright 1997 by the Education Commission of the States (ECS) and the National Center for Research on Evaluation, Standards and Student Testing (CRESST). All rights reserved.

Postage and handling charges if your order totals: Up to $10.00, **$3.00**; $10.01-$25.00, **$4.25**; $25.01-$50.00, **$5.75**; $50.01-$75.00, **$8.50**; $75.01-$100.00, **$10.00**; over $100.00, **$12.00**.

Generous discounts are available for bulk orders of single publications. They are: 10-24 copies, 10% discount; 25-49 copies, 20% discount; 50+ copies, 30% discount.

This work is supported by a grant to the Education Commission of the States by the National Science Foundation, grant number REC-9154539, Jane Armstrong, Principal Investigator. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The work reported in this publication also was supported under the Educational Research and Development Center Program PR/Award number R305B600002, as administered by the Office of Educational Research and Improvement, U.S. Department of Education. The findings and opinions expressed in this publication do not reflect the position or policies of the National Institute on Student Achievement, the Office of Educational Research and Improvement or the U.S. Department of Education.

A technical version of this paper, *Standards-Led Assessment: Technical and Policy Issues in Measuring School and Student Progress* is available from CRESST, 10920 Wilshire Blvd. Suite 900, Los Angeles, CA 90024-6511 or through their web site at http://www.cse.ucla.edu.

A Policymaker's Guide to Standards-Led Assessment

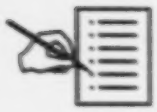
TABLE OF CONTENTS

EXECUTIVE SUMMARY iv

1. INTRODUCTION AND BACKGROUND 1
   - Content and Performance Standards 1
   - The Role of Assessment in Standards-Led Reform 2
   - What's Different About Standards-Led Assessments? 3
   - Assuring Quality of Assessments in Standards-Led Systems 6

2. CHALLENGES FOR STANDARDS-LED ASSESSMENT 7
   - Building State and Local Consensus 7
   - Do the Standards Provide a Solid Foundation for Assessment and Alignment? 8
   - Are Standards Aligned With Assessment and Instruction? 8
   - Defining Different Levels of Performance 9
   - Assuring Accurate Measures in Standards-Led Systems 10
   - What Does Progress Mean? 10
   - Accountability and Stakes: How Are Scores Reported and Used? 10
   - What Does *All* Students Mean? 11
   - What About Costs? 14
   - Legal Defensibility and High-Stakes Student Certification 14
   - Support and Challenges for Building Local Capacity 15
   - The Relationship to District and Classroom Assessments 17
   - Conclusion 18

ENDNOTES 19



EXECUTIVE SUMMARY

States across the country are setting tough new standards, defining what students should know and be able to do. To help students meet these standards — and to measure their progress in doing so — many states are also designing and implementing new assessment systems.

Assessments play a pivotal role in standards-led reform, by:

- Communicating the goals that school systems, schools, teachers and students are expected to achieve;
- Providing targets for teaching and learning; and
- Shaping the performance of educators and students.

Coupled with appropriate incentives and/or sanctions — external or self-directed — assessments can motivate students to learn better, teachers to teach better, and schools to be more educationally effective.

What's Different about Standards-Led Assessments?

Unlike more traditional assessments, standards-led assessments are closely linked to curriculum, producing a tight coupling between what is taught and what is tested. Unlike norm-referenced tests, which compare each student's performance to that of others, standards-led assessments incorporate pre-established performance goals. And unlike multiple-choice exams, many standards-led assessments require students to demonstrate a broad range of problem-solving skills — the very skills students will need for future success.

These "authentic" or "performance" assessments typically engage students in real-world problems, rather than artificial exercises. Such assessments not only measure students' ability to master complex tasks but also model those tasks for teachers, providing examples for use in the classroom.

Performance assessments that require extended responses must be scored by expert judges, using clearly specified scoring guides. The development and application of such scoring guides presents teachers with a rare opportunity to discuss new standards and performance expectations. Examining actual responses helps teachers understand the strengths and weaknesses of their students' learning and plan appropriate instructional activities.

What makes for a sound assessment? Two major criteria are typically cited: *validity*, the degree to which particular uses and interpretations of assessment results are justified; and *reliability*, the degree to which scores are free of measurement error. For standards-led assessment, another key is *alignment* — the degree to which the assessment adequately reflects the standards on which it is supposed to be based. An assessment that is mismatched with a given set of standards may undermine learning, by focusing attention on less important skills or knowledge at the expense of others and more important ones.

Challenges for Standards-Led Assessment Systems

- **Building state and local consensus**
  If public opinion polls are any indication, the concept that students should be held to high academic standards enjoys broad support. Experience shows, however, that such support can be fragile. The diversity of opinion on what students should learn and schools should teach makes it imperative to involve the public in the development of standards *and* assessments. Building a broad consensus requires not just a series of public hearings and opportunities for input and review but a comprehensive process that fully involves the public, ensuring that its concerns are understood and addressed.

- **Providing strong standards**
  Achieving consensus on standards that are broad and vague is no challenge — who would disagree that all students must be able to "communicate effectively"? But when standards are stated in such general terms, they offer little help for the students who must meet them or for the teachers and schools attempting to assess student progress. Available evidence suggests that many states' current standards are not strong enough to support rigorous content-based assessment.

- **Aligning standards with assessment and instruction**
  Many states and localities develop standards and assessments at the same time, rather than following the more logical sequence: standards first, assessments second. Indeed, some states patch

together assessment systems using whatever assessments are available, sacrificing the "custom fit" they would gain by developing assessments from scratch. Systems that rely exclusively on multiple-choice exams cannot show how well students are performing on the full range of skills and understandings covered by standards.

Ultimately, classroom curriculum and instruction should also be aligned with standards and assessments. Yet this alignment depends in turn on teachers' ability to understand — and obtain the resources and expertise to help their students meet — the expectations embodied by new assessments. Fairness demands that students not be held accountable for goals they have had an inadequate opportunity to reach.

- **Assuring accurate measures**
  Performance assessments, which ask students to create a response rather than choose one from a list, generally provide a better gauge of complex thinking skills. But scoring such assessments requires more time, usually more money, and consensus among judges on the quality of the response. To furnish a stable estimate of student capability, most assessments now being developed incorporate a broad range of tasks, reflecting the full scope of the standards. When measuring the performance or progress of a school or district rather than an individual student, assessments can also assign different tasks to different samples of students (a practice known as matrix sampling).

- **Defining progress**
  The progress of schools, districts and states is typically defined by the performance of successive cohorts of students: Are more fourth-graders, for example, demonstrating proficiency in math standards this year than last? Federal law requires that states define "adequate yearly progress" in terms of students' performance on the states' standards-led assessment, determine whether their schools are making such progress, and target an "appropriate" date by which all Title I students will perform at either the proficient or advanced level. States must then set an annual rate of improvement that is both "substantial" and "sufficient" to achieve that goal.

- **Setting the stakes**
  What schools do with assessment results — whether simply reporting them, at one end of the spectrum, or making graduation contingent on them, at the other — can have profound effects on students. Assessments also can be used to hold educators and schools accountable for students' performance. Districts may use the results as the basis of explicit rewards (e.g., cash grants) or sanctions (reassignment or dismissal of staff, administrative take-overs).

- **Including all students**
  Standards are designed to raise expectations for *all* students. Including limited English speaking students and students with disabilities in an assessment may require a variety of different accommodation strategies, from the allotment of extra time to the provision of oral assessments or translation to other languages. Students with learning disabilities — who account for the largest group historically excluded from assessments — may be able to complete assessments, in part or in full, without special accommodations. (Those with severe cognitive disabilities may require a separate system of assessments.)

- **Estimating costs**
  While the costs of assessments vary widely, those requiring extended student responses — to be judged by teachers or other subject-matter experts — cost substantially more than multiple-choice tests. Administering a machine-scorable test may cost between $5 and $8 per student; assessments that require a mix of short answers and extended written responses can easily cost two or three times as much. Estimates for more elaborate performance assessments range from $30 to $70 per student.

- **Addressing legal challenges**
  Assessments are most likely to face legal challenges when high stakes — whether to graduate a student, whether to endorse a diploma — are attached to the results. Challenges also can be expected when assessments produce an adverse impact on historically disadvantaged groups: substantially higher failure rates for African American or Hispanic students, for example. Such evidence does not, by itself, establish the unfairness of an assessment or an intent to discriminate. But the identification of an adverse impact can — and often does — trigger a legal challenge.

Among the other most likely triggers:

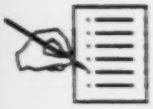
- The "use of processes perceived to be unfair, arbitrary, or capricious"
- The "suggestion that specific attitudes or values are being assessed"
- The "failure to provide all accommodations requested by the disabled"
- The assessment of "knowledge or skills that examinees have not had the opportunity to learn."

The likelihood of legal challenges argues against attaching high stakes to assessment results too soon. Designing a reliable standards-led assessment system is a complex and time-consuming process. It will take just as much time for teachers, schools and students to understand the expectations such a system raises — and to meet them.

- **Building local capacity**
  Research shows that most teachers treat performance assessments seriously and incorporate the underlying goals in their instruction. At the same time, though, many principals and teachers report serious concerns about the demands new assessments place on themselves and their schools. In particular, they report, teachers need time to become familiar with new standards, assessments and administration requirements; to understand how new forms of assessments are developed and scored; to apply criteria for assessing students' work; and to acquire enough information and pedagogical knowledge to change their practices. Providing appropriate resources and sufficient opportunities for professional development is equally important.

- **Distinguishing assessments**
  An assessment that attempts to perform too many functions — student diagnosis, curriculum planning, program evaluation, instructional improvement, accountability, certification, public communication — will inevitably do none well. It is important, therefore, to distinguish appropriate roles for different assessments, at the district, school and classroom level. A cohesive system ensures that teachers and students understand what is important to learn and how well they are doing.

  Teachers routinely use a wide variety of formal and informal assessments to gauge student progress, assign grades, motivate attention, provide feedback and adapt instruction to student needs. Similarly, students regularly engage in self-assessment, as they study and attempt to solve problems and monitor their own progress. Together, all of these assessments provide teachers and students with the detailed understanding and continual feedback they need to guide effective, ongoing learning. It is essential that these assessments reflect state standards.



CHAPTER 1

INTRODUCTION AND BACKGROUND

States across the country are setting tough new standards, defining what students should know and be able to do. To help students meet these standards — and to measure their progress in doing so — many states are also designing and implementing new assessment systems.

These systems hold substantial promise for supporting improved student performance, but their effectiveness turns on a number of factors. This paper lays out the most important such factors, as well as some of the lessons learned over the last decade by states and localities at the center of the assessment debate.

"Improving student performance requires a clear picture of what you want to accomplish, a comprehensive measurement system to gauge progress, and a commitment to act on the results to make appropriate changes."

— Colorado Governor Roy Romer

Standards must be specific enough to enable everyone (students, parents, educators, policymakers, the public) to understand what students need to learn. They also must be precise enough to permit a fair and accurate appraisal of whether the standards have been met. While they do not mandate a particular curriculum, textbook or instructional approach and may be achieved in a variety of ways, standards must make clear what is expected of students.

Content and Performance Standards

States and localities typically distinguish two types of inter-related standards: those that specify the *content* (what students should know or be able to do at different points in their education); and those that specify the *performance* (how well they should be able to do it). Ideally, performance standards indicate the evidence required to demonstrate fulfillment of content standards (e.g., essay, mathematical proof, scientific experiment, project, exam) as well as the quality of performance that will be deemed acceptable (what merits a passing grade or an 'A').[1]

By raising expectations for all students, standards mark an important first step in improving education. But standards alone cannot produce the desired improvement. Curriculum specifications and materials, resource guides, professional development and assessments are equally instrumental. While our focus is limited to assessments, the success of standards-led reform requires a set of systematic changes throughout the educational system.

Examples of Content Standards Statements
Colorado, Geography, Standard 4: **"Students understand how economic, political, cultural, and social processes interact to shape patterns of human populations, interdependence, cooperation, and conflict.... In grades K-4, what students know and are able to do includes ... identifying the causes of human migration"** *(Colorado Model Content Standards for Geography, Colorado Department of Education, adopted June 1995, amended November 1995).*
Missouri, Science Standard: **"In Science, students in Missouri public schools will acquire a solid foundation which includes knowledge of ... properties and principles of force and motion"** *(The Show-Me Standards, Missouri Department of Elementary and Secondary Education, October 1995).*
Oregon, Grades 6-8 Reading Standard and Benchmark: **"Demonstrate inferential comprehension of a variety of printed materials."** *Associated Grade 8 Benchmark:* **"Identify relationships, images patterns or symbols and draw conclusions about their meaning"** *(By Grade Level Common Curriculum Goals, Grades 6-8 Content and Performance Standards, Oregon Department of Education, August 1996).*
Virginia, Grade 5, United States History and Social Science Standards of Learning, Standard 5.3: **"The student will describe colonial America, with emphasis on ... the principal economic and political connections between the colonies and England"** *(Standards for Learning for Virginia Public Schools, Board of Education, Commonwealth of Virginia, June 1995).*

The Role of Assessment in Standards-Led Reform

Assessments play a pivotal role in standards-led reform, by:

- Communicating the goals that school systems, schools, teachers and students are expected to achieve;
- Providing targets for teaching and learning; and
- Shaping the performance of educators and students.

Coupled with appropriate incentives and/or sanctions — external or self-directed — assessments can motivate students to learn better, teachers to teach better, and schools to be more educationally effective.

Assessments communicate goals. All assessments, whether standards-led or not, reveal the expectations of their creators. Students seeking to divine their teachers' wishes often find more clues in past exams than in course syllabi, lectures or reading assignments.[2] Over time, the "tradition of past exams" (as George Madaus describes it) can effectively define the curriculum — especially when students' performance on exams carries important consequences.[3]

Assessments provide targets. Assessments not only elucidate standards, they also provide performance targets for instruction. Assessments focus attention on a particular set of skills and knowledge — those that must be mastered to "meet the standard." Assessments offer operational examples of what students should know or be able to do. They also tell students how good is "good enough," by defining different levels of proficiency.

Assessments shape performance. Standards-led reform hinges on the premise that making expectations explicit will prompt greater effort from both teachers and students — effort focused, by assessments, on specific performance targets. The capacity to motivate and focus effort makes the assessment a powerful tool in the teacher's instructional arsenal. Well-conceived assessments — covering ground that corresponds to course goals and priorities specified in the syllabus — can focus student attention on the knowledge and skills that are deemed most important to learn. Poorly conceived assessments can prompt students to cram soon-to-be-forgotten facts and figures, in the

knowledge that simple regurgitation will secure a passing grade.

The influence of tests on student behavior is not limited to course-based examinations. Large enrollments in preparation courses for the bar exams, the medical boards, and college, graduate and professional school admissions tests also attest to the influence of examinations on students.

Assessments can shape the behavior not only of students, but also of teachers — who often use tests as models for curricular and instructional design. Teachers help students prepare for tests by devoting large shares of class time to test-like activities, especially when the pressure for high scores increases.[4] Over-reliance on multiple-choice exams, in particular, has encouraged teachers to "drill and kill" their students with basic skills worksheets. The result is WYTIWYG ("What You Test Is What You Get") — another reason that well-conceived assessments are so important.

Ultimately, local or state assessments also can shape the behavior of entire school systems. Raising expectations, particularly in the form of high-stakes tests, can prompt authorities to seek the resources their schools or districts need to help students achieve.

What's Different About Standards-Led Assessments?

The role of assessments in increasing accountability and stimulating improvement is not, of course, unique to standards-led reform. What makes standards-led assessment different from their more traditional counterparts?

Closely linked to curriculum. With a few notable exceptions (such as Advanced Placement and New York Regents exams), most externally imposed assessments in the United States measure generic skills and achievements — intentionally decoupled from any specific curriculum, course of study or content standards. Standards-led reform, in contrast, advocates a tight coupling between what is taught and what is tested. The power of assessments to shape teachers' practice, once seen as an unfortunate and unintentional side effect, becomes desirable — indeed, strengthened — as the stakes attached to standards are raised.

Assessments can shape the behavior not only of students, but also of teachers — who often use tests as models for curricular and instructional design.

MULTIPLE CHOICE ITEM

1. *Which of these numbers is less than 50 and greater than 30?*

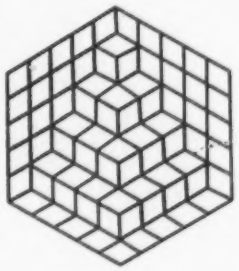
1 10
2 60
3 20
4 40

GENERAL PURPOSE ANSWER SHEETS

	A	B	C	D		A	B	C	D		A	B	C	D
1	①	②	③	④	10	①	②	③	④	20	①	②	③	④
2	①	②	③	④	11	①	②	③	④	21	①	②	③	④
3	①	②	③	④	12	①	②	③	④	22	①	②	③	④

PERFORMANCE TASK

How many cubes were needed to build this figure? Explain several different ways of solving this problem. What patterns do you see? Analyze the patterns. Use correct terminology.



Adapted from: "Better Tests Give a Clearer Picture," Edmonds School District, Lynnwood, Washington.

Students compared to standard of performance, not to other students. Standards-led assessments compare student accomplishment to pre-established performance goals, rather than to the performance of other students. The standard is supposed to be absolute, independent of the proportion of students who meet it. Norm-referenced tests, in contrast, describe what students can do relative to other students: The fact that a student scores at the 60th percentile in math, for examples, tells us only that she fares as well as or better than 60% of her peers — not how much mathematical skill she has mastered.

The current focus on absolute standards mirrors an earlier emphasis on "minimum competency," a reform movement popular in the 1970s and 1980s. Then, as now, reformers sought to improve education by holding educators and students accountable for achieving standards of performance, using tests for high school graduation or grade-to-grade promotion. But in contrast to today's reformers — who emphasize high, rigorous standards — the earlier group targeted only basic skills. And unlike the multi-level standards-led assessments, minimum-competency tests typically employed multiple-choice items on a pass-fail basis.

Incorporate new forms of assessment. Standards-led assessments often take new forms — requiring students, for example, to write an essay, solve a real-life math problem, or design and conduct a hands-on science experiment. Unlike machine-scanned multiple-choice tests, these "performance assessments" (also called "alternative," "authentic" or "direct") are typically scored by humans, who examine student work and apply agreed-upon criteria.

Such assessments capture a broader range of complex thinking and problem-solving skills — skills students will need for future success. The move toward performance assessment also reflects a new emphasis on students' constructive engagement in the learning process. While multiple-choice tests can, if well designed, do more than measure simple recall of discrete facts and isolated basic skills, they sometimes overemphasize such abilities at the expense of more complex reasoning. That tendency can make such tests inadequate, particularly as models of higher-order instruction.

Performance tests emerged in response to a call for "assessments worth teaching to."[3] Lauren and Dan Resnick articulated that demand in three "guidelines for accountability assessments":

- "You get what you assess";
- "You do not get what you do not assess"; and
- "Build assessments toward which you want educators to teach."[4]

To satisfy these aims, assessments need to have a number of features, six of which are summarized in the sidebar. The first, "involving activities that are valued in their own right," is the goal of "authentic assessments" — engaging students in "real world" problems rather than artificial tasks. Other features emphasize assessments' instructional compatibility, value and appropriateness for particular purposes (e.g., accountability or professional development).

Value for professional development. Performance assessments are useful not only for measuring students' ability to master complex tasks but also for modeling those tasks for teachers. Such assessments incorporate the kinds of open-ended, complex problems that students will face in the real world — and simultaneously serve as examples teachers can use in their classrooms.

The need for assessment models should come as no surprise. Many, if not most, teachers were trained at a time when basic skills — rather than today's higher standards — formed the focus of academic achievement. Many current principles of effective teaching and learning were only recently developed. Indeed, assessment itself has long been neglected in teacher preparation.

Performance assessments also serve another function in teachers' professional development. Those that require extended responses must be scored by expert judges, using clearly specified scoring guides. Teachers' participation in the development and application of such scoring guides not only capitalizes on their subject-area expertise, but also provides them with a rare opportunity to discuss new standards and performance expectations. Examining actual responses helps teachers understand the strengths and weaknesses of their students' learning and plan appropriate instructional activities. Indeed,

Desired Features of Assessments

- Assessment tasks should involve activities that are valued in their own right.
- Assessments should model curriculum reform.
- Assessment activities should contribute to instructional improvement by focusing on instruction targets that are consistent with the goals of instructional activities.
- Assessments should provide a mechanism for staff development.
- Assessments should lead to improved learning by engaging students in meaningful activities that are intrinsically motivating.
- Assessments should lead to greater and more appropriate accountability.

Source: From R. L. Linn & E. L. Baker (1996). *Can Performance-Based Assessments be Psychometrically Sound?* In J. B. Baron & D. P. Wolf (Eds.), *Performance-Based Student Assessment: Challenges and Possibilities.* Ninety-fifth Yearbook of the National Society for the Study of Education, Part I. (pp. 84-103) Chicago, University of Chicago Press.

BEST COPY AVAILABLE

CRESST Criteria for Evaluating Assessment
Consequences. To what extent are intended positive consequences achieved? What are the unintended negative consequences?
Fairness. Does the assessment enable students, regardless of race, ethnicity, gender or economic status, to show what they know and can do?
Transfer and Generalizability. Will the results of an assessment provide accurate generalization about student achievement?
Cognitive Complexity. Does the assessment require students to pursue complex thinking and problem solving?
Content Quality. Is the assessment content consistent with the best current understanding of the subject matter?
Linguistic Appropriateness. Does the assessment allow students to display what they know and are able to do without being swamped by language demands not required by the content?
Instructional Sensitivity. Does effective instruction produce improvements in performance?
Curricular Importance. How important are the goals measured by the assessments? Do they measure important content standards?
Content Coverage. To what extent is the full range of the key elements of the content standards covered?
Meaningfulness. Do students find the assessment tasks realistic and worthwhile?
Practicality and Cost. Is the information about students worth the cost and time to obtain it?
Source: National Center for Research on Evaluation, Standards, and Student Testing, Los Angeles, UCLA.

many teachers describe their participation in well conceived scoring sessions as one of the most valuable parts of their professional development.

Assuring Quality of Assessment in Standards-Led Systems

What makes for a sound assessment? Two major criteria are typically cited: *validity*, the degree to which particular uses and interpretations of assessment results are justified; and *reliability*, the degree to which scores are free of measurement error.

Recent research has expanded these criteria, as the sidebars on this and the next page show. One list, developed by the National Center for Evaluation, Standards and Student Testing (CRESST), emphasizes the consequences and "fairness" of assessments.[7] The other adopted by the National Council of Teachers of Mathematics (NCTM), includes "equity" and "openness."[8]

Alignment with standards is imperative. How do the criteria for judging a standards-led assessment differ, if at all, from the list of factors used to evaluate any other test? The answer, in a word, is *alignment* — the degree to which the assessment adequately reflects the standards on which it is supposed to be based.

The point seems obvious, but consider the alternative: An assessment that is mismatched with a given set of standards may undermine learning, by focusing attention on less important skills or knowledge at the expense of others. If standards emphasize critical thinking and explanation in history, for example, then an assessment aligned with the standards must require students to think critically about historical information and explain historical trends or events. A misaligned assessment that simply asked students to recall dates and names would thwart the original intent.

While no assessment can capture the full range of content standards in a given discipline, some may stretch further than others. As noted earlier, multiple-choice tests alone are likely to shortchange skills and knowledge that are harder — but no less important — to assess. The discussion in the next chapter points to the need for a mix of assessment formats.

CHAPTER 2
CHALLENGES FOR STANDARDS-LED ASSESSMENT

Beyond the technical concerns, designing an effective assessment system requires enormous leadership and careful planning. Building state and local consensus and ensuring a sufficient degree of alignment between assessment and instruction — among other challenges — are key to this process.

Building State and Local Consensus

If public opinion polls are any indication, the concept that students should be held to high academic standards enjoys broad support. Experience shows, however, that such support can be fragile. "The consensus breaks down ... in moving beyond this belief in the need for standards and assessment to questions about what those standards should be and how students should be taught and tested."[9] At the national level, for example, the warm reception accorded the NCTM Curriculum and Evaluation Standards[10] has not been extended to standards in other content areas such as history and English-language arts.[11] Indeed, moves to set standards in these areas have prompted fierce debates.

Such dissension is hardly surprising. The very specificity that makes standards and assessments valuable educational targets — by defining what is important for students to know — makes them targets of criticism as well. Moreover, "standards involve much more than determinations of what knowledge is of most worth; they also involve social and cultural differences, and they frequently serve as symbols and surrogates for those differences."[12] The cancellation of the California Learning Assessment System (CLAS), for example, was largely the result of organized opposition to (1) the content emphases that gave little attention to basic skills such as phonics, spelling and arithmetic facts; and (2) assessment exercises that opponents claimed "promoted inappropriate values such as violence and the questioning of authority."[13]

CLAS's active opponents, like the critics of standards and assessments in other states, represent a relatively small, albeit vocal, minority. The questions these critics raise, however, may reflect broader concerns: "[R]ecent surveys about teaching of mathematics and writing point to fundamental differences between the curricular values of education reformers and large segments of the public."[14]

The diversity of opinion on what students should learn and schools should teach makes it imperative to involve the public in the development of standards *and*

NCTM Assessment Standards for School Mathematics

The mathematics standard: "Assessment should reflect the mathematics that all students need to know and be able to do."

The learning standard: "Assessment should enhance mathematics learning."

The equity standard: "Assessment should promote equity."

The openness standard: "Assessment should be an open process."

The inferences standard: "Assessment should promote valid inferences about mathematics learning."

The coherence standard: "Assessment should be a coherent process."

Source: National Council of Teachers of Mathematics, *Assessment Standards for School Mathematics*, Reston, VA, 1995.

BEST COPY AVAILABLE

assessments. While educational reformers may bring strong beliefs in "constructivist" learning theories, other more traditional skills — spelling, phonics, multiplication tables, knowledge of historical dates and locations — also must be considered. Ignoring the wishes of sizable segments of the public will jeopardize the entire enterprise.

Building a broad consensus requires not just a series of public hearings and opportunities for input and review but a comprehensive process that fully involves the public, ensuring that its concerns are understood and addressed.

The experiences of California, Kentucky and North Carolina highlight the need for strong political leadership — especially, as Lorraine McDonnell has found, when the introduction of standards-led reform entails major changes:

> Fundamentally different approaches to teaching and testing need articulate spokespersons who firmly believe in the ideas and who can persuade parents and the general public that these strategies will produce positive gains for individual students and for the state as a whole. That support has to come from people who are visible and whom the public feel can be held responsible for the outcomes. For that reason, the support needs to come [from] people who are electorally accountable, and not just from professional educators and unelected officials within the education establishment.[15]

Do the Standards Provide a Solid Foundation for Assessment and Alignment?

Achieving consensus on standards that are broad and vague is no challenge — who would disagree that all students must be able to "communicate effectively"? But when standards are stated in such general terms, they offer little help for the students who must meet them or for the teachers and schools attempting to assess student progress.

Consider, for example, a standard that requires students to "understand ideas and documents within historical contexts." A variety of multiple-choice, short answer or essay questions might be appropriate assessment tools. But which ideas or documents should students understand? In what context? And what constitutes adequate evidence of understanding?

Since the standard itself provides few clues, the answer must come from those who develop the assessment — ideally, through discussion with educators, students, parents and the public. Once an adequate assessment has been developed, some criteria must be created to score student responses. Even explicit standards leave considerable room to specify assessment tasks and to distinguish levels of student performance.

To be effective, standards must be "written in clear, explicit language ... firmly rooted in the content of the subject area, and ... detailed enough to provide significant guidance to teachers, curriculum and assessment developers, parents, students and others who will be using them."[16] Yet available evidence suggests that many states' current standards are not strong enough to support rigorous content-based assessment. A lack of alignment threatens the success of the states' reform efforts.

Are Standards Aligned With Assessment and Instruction?

Alignment is easier said than done. The difficulty stems, in part, from states' and localities' decisions to develop standards and assessments at the same time, rather than following the more logical sequence: standards first, assessments second. This decision — sometimes the result of budgetary or scheduling pressures — makes alignment more cumbersome. Indeed, some states patch together systems using whatever assessments are available, sacrificing the "custom fit" they would gain by developing assessments from scratch. (There are, of course, notable exceptions.)

The result is often a superficial correspondence between standards and assessments. Both may cover the same topics, but fall short of alignment in other respects.[17] Do the assessments and standards reflect the central concepts and enduring themes of the discipline? Do the assessment tasks call for the kinds of complex thinking and problem-solving capabilities specified by the standards? Are the types of problem situations similar and equally authentic?

Few systems can meet this test. While almost all states and many districts have developed (or are developing) high standards for student performance,[18] many assessment programs still rely exclusively on standardized multiple-choice exams.[19] The data these programs produce can tell schools, parents or the public

how well students are performing only on some aspects of the standards, not on the full range of skills and understandings covered by the standards.

Ultimately, classroom curriculum and instruction should be aligned with standards and assessments. In the absence of such alignment, students cannot acquire the knowledge and skills they need to achieve the standards. Yet this alignment depends in turn on teachers' ability to understand — and obtain the resources and expertise to help their students meet — the expectations embodied by new assessments. Fairness demands that students not be held accountable for goals they have had an inadequate opportunity to reach.

Defining Different Levels of Performance

Content and performance standards articulate what students must know and be able to do to show that they have attained the learning they will need for future success. But a desire to chart students' progress in greater detail has led many states to define not only "what's good enough" but several other levels of performance as well.

The Improving America's Schools Act (IASA) of 1994 requires that states adopt three levels of performance: "proficient," "advanced" and "partially proficient." The proficient level indicates that a student has met the content standards; the advanced level indicates that a student has exceeded them. Lower-performing students are designated as partially proficient.[20] Title I accountability requirements are intended to help all students achieve — or demonstrate adequate annual progress toward — proficiency.

Student performance levels are typically determined through a process of public consensus. Panels of teachers, parents, students and members of the business community are convened to review actual student work, reflecting a range of abilities, and to determine the level at which that work should be classified. Panel members typically make judgments individually, discuss their rationales, and then — aided by statistical programs that convert their judgments into proposed scores — consider the implications in light of actual performance data.[21] Based on this process, a series of "cut scores" is established, allowing student performance on the assessments to be converted into a proficiency level.

Ultimately, classroom curriculum and instruction should be aligned with standards and assessments.

Assuring Accurate Measures in Standards-Led Systems

While asking students to create a response — rather than choose one from a list — may provide a better gauge of complex thinking skills, these performance assessments present special challenges to those who score them. The problem is not the introduction of "subjective" human judgment; humans are involved in multiple-choice tests as well, creating the tests if not scoring them. But scoring performance assessments requires more time, usually more money — and, often trickiest of all, consensus among judges on the quality of the response.

Beyond assuring consensus on scores, how can we be sure results are fair and accurate? The answer depends, in part, on the design of the assessments.[22] There should be enough items to get a stable estimate of student capability. Most assessments now being developed thus incorporate a broad range of tasks, reflecting the full scope of the standards. (Some of these tasks require only a short amount of time to administer; others require considerably more.)

When measuring the performance or progress of a school or district rather than an individual student, assessments can assign different tasks to different samples of students. This approach, known as matrix sampling, ensures comprehensive coverage while minimizing the time each student is required to spend taking the assessment. (Note that matrix sampling is generally not appropriate for assessing individuals' performance.)

What Does Progress Mean?

Assessments are designed to measure students' educational progress — either as individuals or as members of larger groups. The progress of schools, districts and states is typically defined by the performance of successive cohorts of students: Are more fourth-graders, for example, demonstrating proficiency in math standards this year than last? Changes in the distribution of other scores are equally important: What share of students has moved from "proficient" to "advanced"? How many remain "partially proficient"?

What constitutes reasonable progress? Federal Title I programs require that states define "adequate yearly progress" in terms of students' performance on the states' standards-led assessment and determine whether their schools are making such progress. To do so, states must target an "appropriate" date by which all Title I students will perform at either the proficient or advanced level. States must then set an annual rate of improvement that is both "substantial" and "sufficient" to achieve that goal.[23]

To encourage accountability for subgroups of students who are most at risk, the law requires that results be disaggregated and reported separately at the state, district and school level by "gender, each major racial and ethnic group, English-proficient status, migrant status, students with disabilities as compared to students without disabilities, and economically disadvantaged students as compared to students who are not economically disadvantaged."[24]

Kentucky's accountability system represents one approach to these requirements.[25] The state's four categories of performance — "distinguished," "proficient," "apprentice" and "novice" — correspond roughly to the categories specified by the IASA (advanced, proficient, partially proficient, and below partially proficient). The categories carry scores of 140, 100, 40 and 0, respectively.

Kentucky aims to help each school average a score of 100 — possible if all students performed at the proficient level, or, say, if 50% were proficient, 30% were distinguished, and 20% were apprentice (.5x100 + .3x140 + .2x40 = 100). Under this formula, schools can achieve progress not only by increasing the percentage of "distinguished" students but also by reducing the share of "novices." Kentucky's system illustrates the range of factors that might shape other states' definitions of adequate yearly progress.

Whatever definition a state chooses, performance standards and assessments must be comparable from one year to the next. A proficient score in 1999, in other words, needs to represent the same level of skill (in a given area) as a proficient score in 1998 or 1997. Maintaining such consistency requires considerable attention to the technical design of assessments: the number of tasks, task sampling, reuse of tasks and the like.

Accountability and Stakes: How Are Scores Reported and Used?

What schools do with assessment results can have profound effects on students. At a minimum, reporting a student's performance to his or her parents focuses

their attention on their child's educational progress (or lack thereof). Some schools attach higher stakes to assessment results — requiring remedial work, for example, from students who fail to meet a specified standard. Ultimately, a student's graduation or promotion from one grade to another may hinge on his or her performance. (Some schools also record a student's assessment results on an "endorsed diploma.")

Equally high stakes can be applied to educators. Kentucky, for example, reports the assessment results of entire schools rather than individual students. Simply printing such results in the newspaper can increase educators' accountability. Changes in a school's performance also can be used as the basis of more explicit rewards (e.g., cash grants) or sanctions (reassignment or dismissal of staff, administrative take-overs).

As noted earlier, schools receiving Title I funds must demonstrate "adequate yearly progress" in student performance.[26] Schools that fall short two years in a row will receive technical assistance. Those that achieve more than adequate yearly progress for three consecutive years will be designated "distinguished schools."

What Does *All* Students Mean?

Standards are designed to raise expectations for *all* students. Excluding large groups of students from state or district assessments (because of disabilities or language barriers) is no longer considered acceptable.

Including all students in an assessment may require different strategies. For some previously excluded groups, little adaptation is necessary — beyond a commitment to inclusion. Some students may need additional time to complete an assessment. (When speed of response is not a relevant consideration, time limits might be relaxed for all students.)

To accommodate students with limited English, assessments can be offered in other languages — allowing native Spanish speakers, for example, to demonstrate proficiency in math. This approach does present several challenges, however. First, while dozens of languages enter American classrooms, few are common enough to make practical the development of alternative assessments. (In most states, Spanish is the only language other than English with large numbers of native speakers.) Second, many students who have oral

Standards and assessments play a central role in President Clinton's 10-point "Call to Action for American Education in the 21st Century."

The President proposes national tests of individual student performance in reading at grade 4 and mathematics at grade 8, tied to the National Assessment of Educational Progress and the Third International Math and Science Study. The President also calls on states to put an end to social promotion — by requiring students to show what they have learned in order to move from grade school to middle school and from middle school to high school, and by ensuring that a high school diploma actually means something.

Finally, the President's plan recommends the use of standards and assessments to hold schools, administrators and educators accountable. In particular, the plan encourages states and districts to use their authority under the reformed Title I program to hold schools accountable for the assistance they receive — by reconstituting chronically failing schools, among other steps.

U.S. Department of Education. *A Call to Action for American Education in the 21st Century.* Washington, DC, 1997.

MARYLAND'S STUDENT ASSESSMENT PROGRAMS

Maryland's Department of Education and Board of Education have developed several student assessments, including the Maryland School Performance Assessment Program (MSPAP), which assesses students in grades 3, 5 and 8 in six core academic areas; and, the Maryland Functional Testing Program (MFTP), which certifies student mastery of basic skills for high school graduation. A commercially available, norm-referenced test in reading, language and mathematics is provided to districts to put district performance in a national perspective.

Also under development are a new high school assessment program, designed to replace the MFTP, and an Independence Mastery Assessment Program for severely handicapped special education students and students in primary grades to profile their strengths and weaknesses.

The Maryland School Performance Assessment Program

The MSPAP emerged from a 1989 report by Maryland's Sondheim Commission, which called for increasing the accountability and performance of the state's schools. First administered in 1991, the MSPAP provides information on school performance in reading, writing, language usage, mathematics, science and social studies.

To reduce testing time, the assessment is matrix sampled: students are assigned only portions of each content area. School districts are required to administer the California Test of Basic Skills (CTBS)/5 to at least a small sample of their students.

Maryland Public and Teacher Involvement

The MSPAP is well-supported by the public because of their long-term involvement. Parents, business leaders, and state and local legislators provided input in establishing content and performance standards for the MFTP and MSPAP, in the selection of CTBS/5 for the norm-referenced testing program, and will continue to provide input on content and performance standards and the design of the new high school assessment program.

Teacher involvement is the cornerstone of the ongoing success of MSPAP. Teachers develop MSPAP assessment tasks following state design specifications, score MSPAP tests in four regional centers managed by an outside contractor, and helped set MSPAP content and performance standards. Teachers and other local educators have been involved in developing content standards for high schools, the design phase of the new high school assessment program, and will participate in test development, scoring, and setting of performance standards for the high school exams.

Reporting and Using Results

Scores from MFTP, MSPAP and CTBS/5 are available in school and school system report cards, called the Maryland School Performance Reports. Test scores and other information (e.g., dropout and attendance rates) identify declining, low-performing, and improving schools. Declining schools may become eligible for reconstitution and undertake a rigorous school improvement process. Some local school systems place other declining schools on alert. Schools showing the greatest improvement rates receive School Performance Recognition Awards, in the form of funds to support continuing improvement efforts.

In the 1995-96 academic year, all but five of Maryland's 24 school systems scored higher than the year before. The school systems posted gains in 15 out of 18 content areas. The past six years have also seen steady improvements in student dropout and attendance rates.

NEXT CENTURY ASSESSMENT FOR NORTH CAROLINA

In 1993, North Carolina's General Assembly formed a commission to develop a fair and valid assessment system that would measure students' knowledge in real-world terms and provide greater feedback to schools and teachers. The result: the Next Century Assessment for North Carolina. The commission's proposal is based on four principles:

1. *A good accountability system does more than audit student performance; it improves performance.*
2. *Assessment must be credible and open if genuine reform is to occur.*
3. *"Trust but verify" must be the motto of an effective assessment system.*
4. *An effective assessment system must build local capacity to perform high-quality assessment, rather than test externally once a year.*

The Next Century Assessment will use standardized tests, performance-based tasks and portfolios or collections of student work to promote accountability and provide diagnostic and achievement information for individual students. Testing proposed in grades 4, 8, 10 and 12 will be supplemented by a comprehensive examination taken between grades 10 and 12, as well as a graduation project requiring extensive reading, writing and an oral presentation.

The new assessment system extends Accountability in the Basics with Local Control (ABCs), North Carolina's current standardized assessment. Key components of the new system include:

- *State performance tasks, requiring effective application of state standards;*
- *Various oversight mechanisms to ensure that local standards cohere with state standards and that local scoring is reliable; and*
- *A portfolio of student work (a collection of achievement evidence, including state test scores, local work and state performance task results) to be scored locally against state standards.*

Approximately 75 performance tasks and scoring guides will be available through the World Wide Web for use as instructional and assessment tools. Teachers will be required to assess all of their students each year, using tasks selected from this database. A common performance task will be required of all students in the state, in order to calibrate teacher scoring of student work against state standards. Teams of educators from each district will score student portfolios every fall; the results will inform classroom instruction. State assessors will rescore a sample of the portfolios from each district to ensure consistency in scoring.

The Next Century Assessment forms part of a larger accountability effort in North Carolina that includes the establishment of high standards; the creation of a system for basing promotion, retention and graduation decisions on actual student performance (thus ending social promotion); and revised graduation requirements. The state's Education Standards and Accountability Commission is now developing a plan to phase in its recommendations, including guidelines for professional development and teacher education, over the next four years. To date, the State Board of Education has directed the Superintendent of Public Instruction to identify grade levels to serve as benchmark years where students must meet the state standards to be placed at the next grade or level of study. Also underway in Spring 1997 are the field tests for the high school comprehensive exam and the core knowledge exam.

BEST COPY AVAILABLE

proficiency in a first language other than English may not have had formal instruction in that language — and may not, therefore, be able to take a written assessment in their native tongue. For such students, an oral assessment may be necessary.

Students with disabilities may also require special accommodations. Those with visual impairments, for example, may need assessments written in large print or in Braille. Some students may need help recording their responses.

Those with learning disabilities account for the largest group of students historically excluded from assessments. Many such students, who receive individual education plans (IEPs), may be able to complete assessments, in part or in full, without special accommodations. Others may need shorter assessments, more time to complete tasks, oral instructions or oral responses.[27] A tiny fraction (perhaps 0.5% of all students) — those with severe cognitive disabilities — may require a separate system of assessments, dictated by their IEPs.

What About Costs?

How much do standards-led assessments cost? Dependable estimates are difficult to obtain, in part because many of the costs associated with assessment — the time spent by teachers in preparation, administration and scoring — are typically absorbed by schools' normal operations and not priced in a separate budget. The costs of assessments vary widely, depending on the number and length of responses to be judged, the number of judges or scorers, the number of content areas assessed, the number and nature of reports to be produced, and the inclusion of "practice assessments" and other preparation materials (if any).

It is clear, however, that assessments requiring extended student responses — to be judged by teachers or other subject-matter experts — usually cost more than multiple-choice tests, which can be scored by machines. Administering a machine-scorable test may cost between $5 and $8 per student, varying with the volume of tests and the range of scoring services ordered. (That price normally covers individual student score reports, classroom reports and school reports in five or more content areas, as well as subscores in some content areas.) Schools often cut costs by reusing booklets and ordering only answer sheets and scoring services after the first year of administration.[28]

Assessments that require a mix of short answers and extended written responses can easily cost two or three times as much as machine-scorable tests. The New Standards Project reference exams offered by Harcourt Brace Educational Measurement, for example, cost approximately $22 per student (including assessment booklets, basic scoring services and a standard report package for assessments in mathematics or English/language arts).[29]

None of the above estimates includes operational costs for schools, districts or states. And the costs of more elaborate performance assessments — involving, for example, hands-on science tasks — are substantially higher; estimates range from $30 to $70 per student.[30] (Single-subject Advanced Placement tests, by comparison, cost $73 per student, of which $7 is normally returned to the school. Most of these tests include both a multiple-choice section and a section requiring extended student responses.)

Legal Defensibility and High-Stakes Student Certification

Assessments may face a variety of legal challenges. Such challenges are most likely to come when high stakes — whether to graduate a student, whether to endorse a diploma — are attached to assessment results.

Challenges also can be expected when assessments produce an adverse impact on historically disadvantaged groups: substantially higher failure rates for African American or Hispanic students, for example. Such evidence does not, by itself, establish the unfairness of an assessment or an intent to discriminate. But the identification of an adverse impact can — and often does — trigger a legal challenge.[31]

Among the other most likely triggers (according to lawyer and measurement expert Susan Phillips) are:

- The "use of processes perceived to be unfair, arbitrary, or capricious";
- The "suggestion that specific attitudes or values are being assessed";

- The "failure to provide all accommodations requested by the disabled"; and
- The assessment of "knowledge or skills that examinees have not had the opportunity to learn."[32]

The last two challenges — accommodating disabilities and ensuring an adequate opportunity to learn — have proven the trickiest. The Americans with Disabilities Act of 1990[33] requires that disabled students be provided with reasonable accommodations. "The courts have clearly indicated that reasonable accommodations must compensate for aspects of the disability that are incidental to the skill being measured but that test administrators are not required to change the skill being measured to accommodate a disabled examinee."[34] But determining which aspects of a disability are incidental to the skill being measured and what accommodations would alter the nature of that skill is no easy task.

Arguments involving the "opportunity to learn" (OTL) have arisen in prior court cases (including *Debra P. vs. Turlington*, a Florida case challenging the state's minimum competency requirement).[35] Such arguments also are likely to form part of any challenge triggered by evidence of adverse impact: Racial differences in assessment results may reflect disparities in students' opportunities to learn.

The debate over OTL eventually led to the inclusion of voluntary "opportunity to learn" standards in the Goals 2000: Educate America Act of 1994.[36] Proponents argued that it was unfair to hold students accountable for meeting performance goals without giving them the instruction to do so. Critics contended that OTL standards would constrain local practice. Any enforcement of these standards seems possible only through further court action.

The likelihood of legal challenges argues against attaching high stakes to assessment results too soon. Designing a reliable standards-led assessment system is a complex and time-consuming process. It will take just as much time for teachers, schools and students to understand the expectations such a system raises — and to meet them.

Support and Challenges for Building Local Capacity

Standards-led reform requires much more than the adoption of goals or assessments. Systemic change of

Assessments that require a mix of short answers and extended written responses can easily cost two or three times as much as machine-scorable tests.

this kind encompasses instructional resources, professional development and classroom practice.

The introduction of standards-led assessments can, however, serve as a catalyst for other reforms. Research in Vermont, Maryland, Arizona, North Carolina and Kentucky showed that most teachers treat performance assessments seriously and incorporate the underlying goals in their instruction.[37]

At the same time, many principals and teachers report serious concerns about the demands new assessments place on themselves and their schools.[38] In particular, they report, teachers need time to become familiar with new standards, assessments and administration requirements; to understand how new forms of assessments are developed and scored; to apply criteria for assessing students' work; and to acquire enough information and pedagogical knowledge to change their practices.

Standards-led reform requires much more than the adoption of goals or assessments. Systemic change of this kind encompasses instructional resources, professional development and classroom practice.

Effecting meaningful changes in teaching practice is neither easy nor cheap. In Lorraine McDonnell's analysis of two states, teachers did not fully understand the demands of state standards and were unable to discern well-aligned classroom activities — despite the provision of professional development and training.[39]

Who should be responsible for professional development in such cases? Who should pay for it? By remaining mute on these questions, most states have pushed responsibility to the local level. The assumption here seems to be that accountability and incentive structures will prompt school districts to supply adequate support.

For those districts with the requisite resources and expertise, such an assumption may be warranted. In many districts, though, support for new materials or professional development does not exist. According to Mary Lee Smith's study of the now-defunct Arizona State Assessment Program (ASAP), the most dramatic progress occurred in schools that were already changing — and probably would have changed anyway.[40] Schools that lacked the will or capacity to change did not benefit from the ASAP program. (One such school was geographically remote and resource-poor; another regarded its students as incapable of reaching higher goals.)[41] In the end, state mandates offer no panacea.

The Relationship to District and Classroom Assessments

An assessment that attempts to perform too many functions — student diagnosis, curriculum planning, program evaluation, instructional improvement, accountability, certification, public communication — will inevitably do nothing well. It is important, therefore, to distinguish appropriate roles for different assessments, at the district, school and classroom level.

At the same time, these assessments must be aligned with one another and with the standards they serve. A cohesive system ensures that teachers and students understand what is important to learn and how well they are doing.

Oregon's Educational Act for the 21st Century is a good example. Under the law, Oregon students in grades 3, 5, 8 and 10 must take a series of statewide uniform tests and local assessments, based on content and performance standards established at each of these grades. The statewide tests include multiple-choice, essay and math problem-solving questions. The local component includes classroom assignments and other, less easily assessed tasks. While these tasks vary from teacher to teacher and from school to school, all students must complete a number of specified types of assignments and achieve a minimal score.

In each content area, Oregon has established a six-point scoring guide for teachers and schools to use in judging student work. The state requires that students achieve the grade 10 standards (in English, mathematics, science, history, among other subjects) to be awarded a *Certificate of Initial Mastery*. Those who meet the grade 12 standards receive a *Certificate of Advanced Mastery*.[17]

New Mexico's state assessment program provides another example. The state requires the same norm-referenced tests for all students at grades 3, 5 and 8; portfolios of students' writing for grades 4, 6 and (optionally) 8; a high-school competency exam; and district-designed reading assessments for grades 1 and 2 — the results of which must be reported to the state.

Utah, to use a final example, administers a standardized norm-referenced test of all students in grades 5, 8 and 11. The state also offers districts a set of criterion-referenced and performance tests to assess student achievement, based on the state framework.[18]

As these examples show, state and local assessment programs can be aligned in a number of ways. Oregon's inclusion of classroom work, as well as the portfolio assessments used in several other states,[44] represent explicit attempts to link state standards and assessments with classroom practice. Such linkage is critical to student achievement.

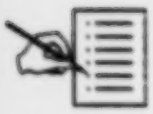
Assessment, of course, is an integral part of the teaching and learning process, occurring continuously in classroom practice. Teachers routinely use a wide variety of formal assessments (exams, pop quizzes, homework assignments, term papers, projects), as well as more informal means (oral questions, class discussion, observation of students' facial expressions), to gauge student progress, assign grades, motivate attention, provide feedback and adapt instruction to student needs.

Similarly, students regularly engage in informal self-assessments, as they study and attempt to solve problems, monitor their own progress and improve their learning. Indeed, teachers and students spend far more time engaged in self-assessment than in completing external tests. Self-assessment also exerts more influence on the day-to-day instructional decisions of teachers and the learning experiences of students. Classroom practice and self-assessment provide teachers and students with the detailed understanding and continual feedback they need to guide effective, ongoing learning.

In light of their pivotal role, it is important that classroom assessment practice and student self-assessments be guided by the same standards on which other assessments are based. External assessments can help in this regard, both by serving as models and by helping teachers understand new standards of student performance. (Oregon's scoring guides represent just such tools.) Sustained support of professional development is equally important.

Conclusion

"A clear picture of what you want to accomplish, a comprehensive measurement system to gauge progress, and a commitment to act on the results to make appropriate changes" — those were Governor Romer's requirements for improving student performance.[45] Content and performance standards are intended to provide the "clear picture" of what needs to be accomplished. Sound assessments aligned with those standards form the "measurement system." Demonstrating the "commitment to act," by providing high-quality instructional resources and extensive professional development, by engaging all students, and by securing broad public support and involvement — that is the challenge which remains.



ENDNOTES

[1]National Education Goals Panel. *Report of Goals 3 and 4 Technical Planning Group on the Review of Education Standards*. Washington, D.C., 1993.

[2]G. F. Madaus. "The Influence of Testing on the Curriculum." In L. N. Tanner (Ed.), *Critical Issues in Curriculum*. Eighty-Seventh Yearbook of the National Society for the Study of Education, part 1, pp. 83-121. Chicago: University of Chicago Press, 1988.

[3]Ibid., p. 93.

[4]See, for example, Corbett & Wilson, 1991; Dorr-Bremme & Herman, 1986; Herman & Golan, 1991; Kellaghan & Madaus, 1991; Shepard, 1991; Smith & Rottenberg, 1991.

[5]L. B. Resnick & D. P. Resnick. "Assessing the Thinking Curriculum: New Tools for Educational Reform." In B. R. Gifford & M. C. O'Connor (Eds.), *Changing Assessments: Alternative Views of Aptitude, Achievement and Instruction* (pp. 37-75). Boston: Kluwer Academic Publishers, 1992, p. 59.

[6]Resnick & Resnick, 1992, op cit., p. 59.

[7]CRESST Web Site: http://www.cse.ucla.edu. See also, R. L. Linn, E. L. Baker, & S. B. Dunbar, 1991. "Complex Performance-Based Assessment: Expectations and Validation Criteria." *Educational Researcher*, 20, 15-21, for a discussion of the rationale for the CRESST criteria and J. L. Herman, P. R. Aschbacher, & L. Winters (1992). *A Practical Guide to Alternative Assessment*, Alexandria, VA: Association of Supervision and Curriculum Development for examples of their application.

[8]National Council of Teachers of Mathematics (NCTM). *Assessment Standards for School Mathematics*. Reston, VA: NCTM, 1995.

[9]L. McDonnell. *The Politics of State Testing: Implementing New Student Assessments* (Technical Report). Los Angeles: UCLA Center for the Evaluation, Standards, and Student Testing, 1996, p. 31.

[10]National Council of Teachers of Mathematics. *Curriculum and Evaluation Standards for School Mathematics*. Reston, VA: NCTM, 1989.

[11]See, for example, K. Diegmueller, "Panel Unveils Standards for History: Release Comes Amid Outcries of Imbalance." *Education Week*, 14, No. 9, November 2, 1994, pp. 1, 10.

[12]L. A. Cremin. *Popular Education and it Discontents*. New York: Harper & Row, 1989, p. 9.

[13]McDonnell, 1996, op cit., p. 34.

[14]Ibid., p. 31.

[15]McDonnell, op cit., p. 68.

[16]American Federation of Teachers. *Making Standards Matter*. Washington, DC: AFT, 1996, p. 19.

[17]N. Webb. *Determining Alignment of Expectations and Assessments in Mathematics and Science Education*. National Institute for Science Education (NISE) Brief, 1(2), January 1997.

[18]AFT, 1996, op cit.

[19]L.A. Bond, D. Braskamp and E. Roeber. *The Status Report of the Assessment Programs in the United States*. Oakbrook, IL: NCREL/Council of Chief State School Officers, 1996.

[20]Public Law 102-382.

[21]See, for example, Wiley. *The New Standards Reference Examination Standards-Referenced Scoring System*. Los Angeles: UCLA Center for the Study of Evaluation (forthcoming).

[22]Available evidence suggests from 5-20 tasks are needed to get a reliable estimate. See, for example, Baker, 1994; Dunbar, Koretz & Hoover, 1991; Linn, Burton, DeStefano, & Hanson, 1995; Shavelson, Baxter, & Pine, 1991; Shavelson, Baxter & Gao, 1993; Shavelson, Mayberry, Li, & Webb, 1990.

[23]Public Law 102-382.

[24]Ibid.

[25]See, for example, Kentucky Department of Education (KDE). *Kentucky Instructional Results Information System: 1992-93 Technical Report*. Frankfort, KY: KDE, 1994.

[26]Public Law 103-382.

[27]See, for example, National Academy of Education (NCE). *Quality and Utility of the 1994 Trial State Assessment in Reading*. Stanford, CA: Stanford University, NCE (especially, chapter 4), 1996.

[28]The reuse of test forms can present concerns if the stakes for teachers or schools are high. See, for example, Linn, R. L. *Assessment-Based Reform: Challenges to Educational Measurement*. William H. Angoff Memorial Lecture Series. Princeton, NJ: Educational Testing Service, 1995.

[29]Harcourt Brace Educational Measurement. *Catalog: Tests and Related Services*. San Antonio, 1997. New Standards Project partner states and districts currently receive a discount on the cost per student.

[30]See, for example, L. McDonnell. *Policymakers' Views of Student Assessment* (Technical Report No. 378). Los Angeles: UCLA Center for the Evaluation, Standards, and Student Testing, 1994, pp. 2-24.

[31]S. E. Phillips. Legal Defensibility of Standards: Issues and Policy Perspectives. In *Proceedings of the Joint Conference on Standard Setting for Large-Scale Assessments of the National Assessment Governing Board (NAGB) and the National Center for Education Statistics (NCES)* (pp. 379-393). Volume II. Washington, DC: NAGB and NCES, 1995.

[32]Ibid., p. 380.

[33]Public Law 101-336.

[34]S. E. Phillips, op cit., p. 391.

[35]Debra P. v. Turlington, 474 F. Supp. 244 (M. D. Fla. 1979), 644 F.2d 397 (5th Cir. 1981); 564 F. Supp. 177 (M.D. Fla. 1983), 730 f.2d 1405 (11th Cir. 1984).

[36]Public Law 103-227.

[37]See, for example, D.M. Koretz, K.J. Mitchell, S. Barron and Keith. *Perceived Effects of the Maryland State Assessment Program* (CSE Technical Report #406). Los Angeles: UCLA Center for the Study of Evaluation, 1996.

D.M. Koretz, S. Baron, K.J. Mitchell and B.M. Stecher. *Perceived Effects of the Kentucky Instructional Results Information System (KIRIS)*. Santa Monica, CA: RAND, 1996.

D. Koretz, B. Stecher, S. Klein, D. McCaffrey and E. Deibert. *Can Portfolios Assess Student Performance and Influence Instruction? The 1991-92 Vermont Experience* (CSE Technical Report No. 371). Los Angeles: UCLA, Center on Evaluation, Standards and Student Testing.

L.M. McDonnell and C. Choisser. *Testing and Teaching: Local Implementation of New State Assessments*. Los Angeles: UCLA Center for the Study of Evaluation (forthcoming).

M.L. Smith. *Reforming Schools by Reforming Assessment: Consequences of the Arizona Student Assessment Program* (CSE Technical Report). Los Angeles: CRESST, 1996.

[38]P.R. Aschbacher. *Issues in Innovative Assessment for Classroom Practice: Barriers and Facilitators* (Technical Report No. 359). Los Angeles: UCLA, CRESST, 1993.

S.A. Wolf and M. Gearhart. *Writing What You Read: Assessment as a Learning Event* (Technical Report No. 358). Los Angeles: UCLA, CRESST, 1993.

[39]See McDonnell, L.M. & Choisser, C. (forthcoming) *Testing and Teaching: Local Implementation of New State Assessments*. Los Angeles: UCLA Center for the Study of Evaluation.

[40]See Smith, M.L. *Reforming Schools by Reforming Assessment: Consequences of the Arizona Student Assessment Program* (CSE Technical Report). Los Angeles: CRESST, 1996.

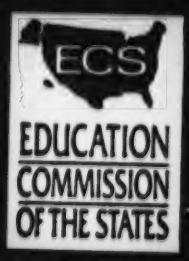
[41]See McLaughlin, M.W. "Learning from Experience: Lessons from Policy Implementation." *Educational Evaluation and Policy Analysis*. 9 (2), 1987, (pp. 171-178); McDonnell & Choisser, op cit.

[42]This description focuses on requirements for the Certificate of Initial Mastery, since those for the Certificate of Advanced Mastery as still under development. See *Adopted Common Curriculum Goals: Content and Performance Standards and Scoring Guides*. Oregon Department of Education, October 1996.

[43]For additional examples and more detail, see Bond, L.A., Braskamp, D. & Roeber, E. *State Student Assessment Programs Database School Year 1994-1995*. Oakbrook, IL: North Central Regional Educational Laboratory/Council of Chief State School Officers, 1995.

[44]A portfolio is a collection of student work designed to show progress over time and level of accomplishment.

[45]Governor Romer's statement quoted by Colorado Education Goals Panel. *Partnerships for Educating Colorado Students: Bringing Out the Best in All of Our Students*. 1995.



Education Commission of the States
707 17th Street, Suite 2700
Denver, Colorado 80202-3427

BEST COPY AVAILABLE